



**UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION III
Four Penn Center
1600 John F Kennedy Blvd
Philadelphia, Pennsylvania 19103-2852**

June 1, 2022

Mindy S. Neil, Environmental Resources Program Manager
Division of Water and Waste Management
West Virginia Department of Environmental Protection
601 57th Street SE
Charleston, West Virginia 25304

Dear Ms. Neil:

Thank you for the opportunity to review and provide comments during the public comment period on the West Virginia Department of Environmental Protection's (WVDEP) Draft 2018-2020-2022 West Virginia Integrated Water Quality Monitoring and Assessment Report (draft IR), which was made available for public comment on April 11, 2022 through June 1, 2022. The U.S. Environmental Protection Agency, Region III (EPA) provides the following comments on the draft IR listed below.

Data Management

- EPA commends WVDEP for creating an online StoryMap to facilitate a greater degree of data interaction. It is a useful tool for visualizing and understanding assessment results.
- EPA appreciates WVDEP's efforts to re-delineate assessment units to better fit EPA's ATTAINS data rules, improve consistency and tracking between cycles, better align with TMDL model predictions of impairment and attainment, and improve spatial representation of existing streams in the assessment database.
- EPA thanks WVDEP for the summary of WVDEP's monitoring program in the draft IR and commends WVDEP for maintaining a robust monitoring program.

Third Party Monitoring and Data

- EPA commends WVDEP for clearly describing how WVDEP considers external data in developing the draft IR, the organizations that submitted data, and the rationale for not using some pH data. Please describe if there were any other external data not used for assessment, along with a scientific rationale for those exclusions, including specific quality assurance concerns, as applicable.
- Certain National Pollutant Discharge Elimination System (NPDES) permittees in West Virginia (WV) are required by the terms of their permits to collect chemical and biological samples upstream and downstream of their discharge to monitor receiving water conditions. Please describe these data and explain how these data were evaluated for the draft IR.

Use Assessment Procedures (303(d) Listing Methodology)

Selenium

- The draft IR states that WV collected 174 selenium fish tissue samples. Certain NPDES permittees are also required by the terms of their permits to collect selenium fish tissue data. Page 11 of the draft IR states, “Water column concentrations were the only data considered for this Integrated Report cycle. In the future, once an assessment protocol is developed and data are available, concentration of selenium in fish tissue may result in listings or delistings of selenium impairment.” EPA requests that WVDEP provide a technical rationale for not using the selenium fish tissue data in assessment decisions. Also, please clarify if instream Discharge Monitoring Report sample data were evaluated as part of the water column concentration assessment and provide a technical rationale for any selenium water column data not used for assessment.

Nutrient Stressor Identification

- EPA suggests that WVDEP update its biological stressor identification protocol to distinguish nutrient stressors from bacteria stressors. The current cause of “organic enrichment” does not clearly identify the pollutant stressors.

Narrative Water Quality Criteria for Support of Aquatic Life – Biological Impairment Data

WVDEP has not yet provided an adequate technical, science-based rationale for not using its full genus-level biological dataset to identify impairments. *See* 40 CFR 130.7(b)(6)(iii). WVDEP first identified biological impairments on its 303(d) list in 2002 using family level data. For the 2004, 2006, 2008, and 2010 303(d) lists, WVDEP used the family level West Virginia Stream Condition Index (WVSCI) to assess its family-level biological data. For the 2012 list, following WV legislative action directing the agency to develop and secure legislative approval of new rules to interpret its biological data against West Virginia’s narrative water quality criterion for support of the aquatic life use, WVDEP did not identify new impairments using biological data. EPA partially disapproved West Virginia’s 2012 303(d) list due to WVDEP’s failure to evaluate its biological data. Because EPA did not wish to preempt WVDEP’s efforts to develop a new methodology for interpreting biological data, EPA added 248 segments to the 303(d) list using WVSCI and the associated family-level data, but noted that EPA expected WVDEP to use its genus level data in subsequent lists. In the same action, EPA also determined that the uncertainty zone historically used by WVDEP was not scientifically supported, and EPA used an impairment threshold equal to the 5th percentile of reference scores as originally calculated in WVSCI development to identify biological impairments. For the 2014 303(d) list, WVDEP resumed its use of WVSCI, but did not evaluate genus-level data and did not provide a reasonable basis for not using the genus-level data. EPA partially disapproved WVDEP’s 2014 303(d) list and added 28 impaired stream segments to the 303(d) list based on failing Genus Level Index of Most Probable Stream Status (GLIMPSS^{1,2}) index scores. For the 2016 cycle, WVDEP again submitted a 303(d) list developed only using the family-level data and index (WVSCI). EPA approved the 2016 list.

¹ Pond GJ, Bailey JE, Lowman BM, Whitman MJ. 2011. The West Virginia GLIMPSS (Genus Level Index of Most Probable Stream Status): A Benthic Macroinvertebrate Index of Biotic Integrity for West Virginia's Wadeable Streams. WVDEP. DOI: [10.13140/RG.2.1.4536.3682](https://doi.org/10.13140/RG.2.1.4536.3682)

² Pond GJ, Bailey JE, Lowman BM, Whitman MJ. 2012. Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams. *Environ Mon Assess* 185: 1515-1540.

For the 2018-2020-2022 303(d) list, WVDEP developed a new biological assessment methodology that uses primarily family-level data and only selectively uses genus-level data under limited circumstances. Under the applicable regulations (40 C.F.R. 130.7(b)(5) & (6)(iii)), it is incumbent upon WVDEP to assemble and evaluate all existing and readily available data and to provide a scientifically sound rationale for not using existing and readily available data to develop its 303(d) list. Based upon review of West Virginia's draft IR, WVDEP has not provided a scientifically sound rationale for limiting its use of existing and readily available genus-level data to make assessment decisions. Specifically, the draft IR lacks a sound, scientific rationale for the following:

1. The selective use of genus level data for a small subset of second samples under limited circumstances;
2. The requirement for a second sample to identify impairments for select streams. (As stated in past EPA 303(d) list decision rationales, one GLIMPSS sample is sufficient to make assessment determinations and use of a zone of uncertainty must be statistically valid.);
3. The single sample impairment threshold;
4. The thresholds used to determine when second samples are required;
5. The impairment thresholds for the second sample;
6. The requirement to collect more data to determine impairment but not to determine attainment; and,
7. The use of genus level data to determine stressors causing biological impairment, but not the impairment itself.

On March 23, 2021, EPA submitted a letter to WVDEP that expressed these technical concerns and recommendations related to WVDEP's bioassessment methodology. Upon review of the draft IR, EPA finds that the concerns and recommendations set forth in that letter, which is attached for reference, remain unaddressed. In the absence of a sound scientific rationale, EPA recommends that WVDEP use its genus-level data to assess whether waters are achieving applicable narrative criteria for the support of aquatic life. EPA's analysis, described more fully below, concludes that:

- A. The draft 2018-2020-2022 303(d) list does not provide a technical, science-based rationale for WVDEP's limited and selective use of genus-level data and other aspects of how it is using its biological data.
- B. A single macroinvertebrate sample is robust and sufficient to make an assessment determination without a zone of uncertainty.
- C. Evaluation of genus level data with GLIMPSS results provides information that is not used when evaluating family level data with WVSCI.
- D. Use of family level data does not use bioassessment information collected from November to March and information about taxa loss that should be used to make assessment decisions.
- E. Analysis at the family level does not use information about taxa richness and taxa loss that is available from genus-level data

EPA Analysis

A. To the extent WVDEP relies upon the rationale it provided for its determination not to use genus-level data to develop its 2016 303(d) list, the GLIMPSS reference dataset is sufficiently robust for use in impairment determinations.

In connection with its 2016 303(d) list, WVDEP expressed concern that the GLIMPSS reference dataset contained insufficient quantities of reference samples for certain ecoregion-season combinations, and that the GLIMPSS dataset therefore didn't provide sufficient "confidence" in the resulting assessment determinations. Since our approval of the 2016 303(d) list, EPA has undertaken a more in-depth review of information relevant to the GLIMPSS reference sites. In addition, since the 2016 303(d) list, WVDEP has collected more data, including identification of additional reference sites. Based on EPA's review and the additional information collected by WVDEP, the GLIMPSS reference dataset is sufficiently robust for use in impairment determinations. While the WVSCI reference dataset includes more samples, the family-level WVSCI index applies statewide from April to October. By contrast, the genus-level GLIMPSS index is applicable statewide and all year round (except November), and is calibrated to specific seasons and ecoregions. This calibration results in more refined impairment thresholds depending on the season and ecoregion in which a sample was collected. Since acting on the 2016 list, EPA's review has led EPA to conclude that consideration of the number of reference sites alone, without additional analysis, does not provide an appropriate measure of "confidence" in a biotic index such as the GLIMPSS. The number and quality of reference sites influences underlying statistical measures which evaluate the accuracy and precision of an index to assess biological condition, and it is those statistical measures that should be considered in determining "confidence" in an index. These include (1) reference site standard deviation and coefficient of variation, (2) revisit data confidence intervals and coefficient of variation, and (3) overall GLIMPSS discrimination (DE) and classification efficiencies (CD). These specific performance measures, described in more detail below, provide confidence that GLIMPSS is of sufficient quality to use in regulatory decisions. Additionally, WVDEP has collected more data which improves previous GLIMPSS index performance. The GLIMPSS reference data set and methodology is sufficient to allow use of a single GLIMPSS score to determine attainment with WV's applicable narrative water quality criteria for support of aquatic life for the purposes of the combined 2018-2020-2022 303(d) list.

1. Reference Site Variation

Indices of biotic integrity, such as the GLIMPSS, are developed based on the premise that samples from streams with minimal anthropogenic disturbance (i.e., reference or least-disturbed sites) define biologic expectation, and the degree of deviation from those reference sites determines if a stream segment is impaired. Therefore, accurately characterizing expected (reference) biological condition is critically important to defining impairment. To characterize reference condition, samples from a subset of reference sites are used to define the entire population (of biological communities) at all reference sites.

As is true for any inference that depends on samples to represent a population, the more variation (range of data) you have in your population, the more samples you need to capture that variation and have confidence in the characterization of the population based on a limited subset of samples. So, while the overall number of samples is important, this alone does not provide sufficient information about the ability of reference biological samples to accurately characterize reference population condition; the underlying data variation must be evaluated as well.

Data variation is mostly commonly evaluated using standard deviation and coefficient of variation of reference site sample scores. Standard deviation (SD) measures how far the average value lies from the mean. Coefficient of variation (CV) measures the ratio of standard deviation to the mean. With low sample variation, as evaluated by those performance measures, there is a high degree of confidence that reference condition is accurately characterized based on the available samples and underlying population variability.

Table 1: Reference site performance measures for ecoregion-seasons in GLIMPSS. WVDEP expressed low confidence in the rows shaded in grey in the 2016 303(d) list.

Ecoregion-Season	# of Reference Sites	Mean	Standard Deviation	Coefficient of Variation (CV%)
Mountain Spring	128	71.2	10.9	15.3
Mountain Summer	181	75.8	11.4	15.1
Mountain Winter	35	77.2	8.0	10.3
Large Mountain Summer	53	72.2	10.6	14.6
Plateau Spring	38	76.2	12.8	16.9
Plateau Summer	44	80.0	11.4	14.2
Plateau Winter	26	81.1	8.1	9.9

As *Table 1* demonstrates, despite having over 3 times the number of reference sites, the Mountain Spring and Mountain Summer have almost the same SD and CV as the ecoregion-seasons in grey as to which WVDEP previously expressed concern that the number of reference samples was “too low to provide confidence in the use of these IBIs”. The variability of reference site scores is comparable and low enough across all GLIMPSS ecoregion-seasons to have high confidence in the GLIMPSS reference site population characterization.

2. Precision and Reproducibility (revisit data confidence intervals and coefficient of variation)

Precision is the ability to repeat an environmental measurement or IBI score. Documenting and achieving high precision is necessary to have high confidence in environmental data, including GLIMPSS. The goal of a biological assessment methodology is to measure biotic response to anthropogenic disturbance rather than natural or sampling variability. In biological sampling, “measurement error is introduced from both natural (e.g., patchiness of habitat and associated macroinvertebrates) and methodological (both field and lab methods) sources of

variability. This measurement error is most commonly estimated using repeat or duplicate samples which are collected on the same day, or within one index period (Pond et al., 2011)¹.” Poor index precision could result from inadequate characterization of reference community (and therefore inadequate characterization of natural versus anthropogenic differences in population expectations). To evaluate precision, the GLIMPSS scores were compared from samples collected in different years from the same reference sites (called revisit sites).

To assess reference revisit precision, an ANOVA test of statistical differences was conducted to determine the within-site sample mean square error (MSE or variance). The square root of the MSE provided sample standard deviation. Then, the distribution of standard deviations from all the sites within an ecoregion-season was used to calculate the 90% confidence interval (CI). This 90% confidence interval is an inference using sample data (revisit sites) to characterize the population; it represents the range of true standard deviation from the mean that may be in the reference population. Low CI values indicate high precision. For GLIMPSS, the values indicate no systematic difference between ecoregion and seasons (“seagions”) (the value for Plateau Summer is due to the overall lower mean of these duplicates). The coefficient of variation was also calculated for re-visit samples (Table 2), where lower values indicate better precision.

Table 2: GLIMPSS precision estimates and statistics from reference site re-visits (Pond et al., 2011¹). (The rows shaded in grey are ecoregions WVDEP expressed low confidence in on the 2016 303(d) list.)

Ecoregion-Season	# of revisits	# Total Reference Sites	MSE	SD	Population mean	One-tailed 90% CI	CV (%)
All combined	30	NA	42.6	6.5	80.5	8.4	8.1
Mountain Spring	8	128	27.8	5.3	76.6	6.8	6.9
Mountain Summer	12	181	43.9	6.6	82.9	8.5	8.0
Plateau Spring	3	38	22.0	4.7	78.8	6.0	6.0
Plateau Summer	7	44	70.1	8.4	81.6	10.7	10.3
WVSCI	39	NA	33.2	5.75	86.8	7.4	6.6

The results of the confidence interval and coefficient of variation for revisit data indicate that despite having fewer number of reference sites, the precision is comparable between the seagions evaluated (and to WVSCI) and sufficient for 303(d) and 305(b) assessment.

3. GLIMPSS Discrimination (DE) and Classification Efficiencies (CE)

Discrimination efficiency (DE) evaluates if the GLIMPSS accurately differentiates stressed and reference conditions. If the discrimination efficiency score is high, that indicates that the GLIMPSS accurately determines stream condition. One reason that DE may score low is if reference condition is inadequately characterized (for example, due to high variability in reference population and/or low number of samples sites). However, all GLIMPSS seagions had high discrimination efficiencies, indicating that all seagion models, including Pl Sp, Pl Su, and MT Su>60, are similarly robust and reliable for 303(d) assessment. Discrimination Efficiencies are shown in Figure 1 and 2 below.

Classification efficiency (CE) evaluates how accurately the GLIMPSS determines attainment and impairment, where a high CE indicates high confidence in characterizing streams. Similar to DE, one reason CE might score low is if reference condition was inadequately characterized. CE for the GLIMPSS was calculated with a separate validation dataset (a random set of reference and stressed sites) that were not used to develop the GLIMPSS. CE is the sum of the number of validation reference sites (scoring above the 5th percentile of reference distribution) and the number of validation stressed sites (scoring below the 5th percentile of reference distribution) divided by the total number of validation sites. The CE for the GLIMPSS (Table 3) indicates it performs with a very high level of confidence (CE 85-95%) and reference population was adequately characterized in all seagions.

Table 3: Comparison of number of reference sites and CE in 2016 and 2020.
(The rows shaded in grey are ecoregions WVDEP expressed low confidence in on the 2016 303(d) list.)

Seagion	# Reference Sites 2016	2016 CE	# Reference Sites by 2020	2020 CE
MT Spring	128	94	246	94.5
MT Summer	181	95	305	96.1
MT Summer >60	53	90.9	65	98
MT Winter	29	NA	51	100
PL Spring	44	94	63	84.7
PL Summer	38	89	44	91.5
PL Winter	18	NA	25	88.5

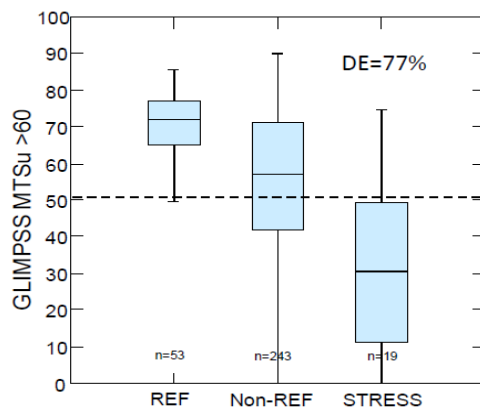


Figure 1. Boxplots of calibration (CAL) GLIMPSS scores between REF, non-REF, and STRESS categories in MT Su >60.

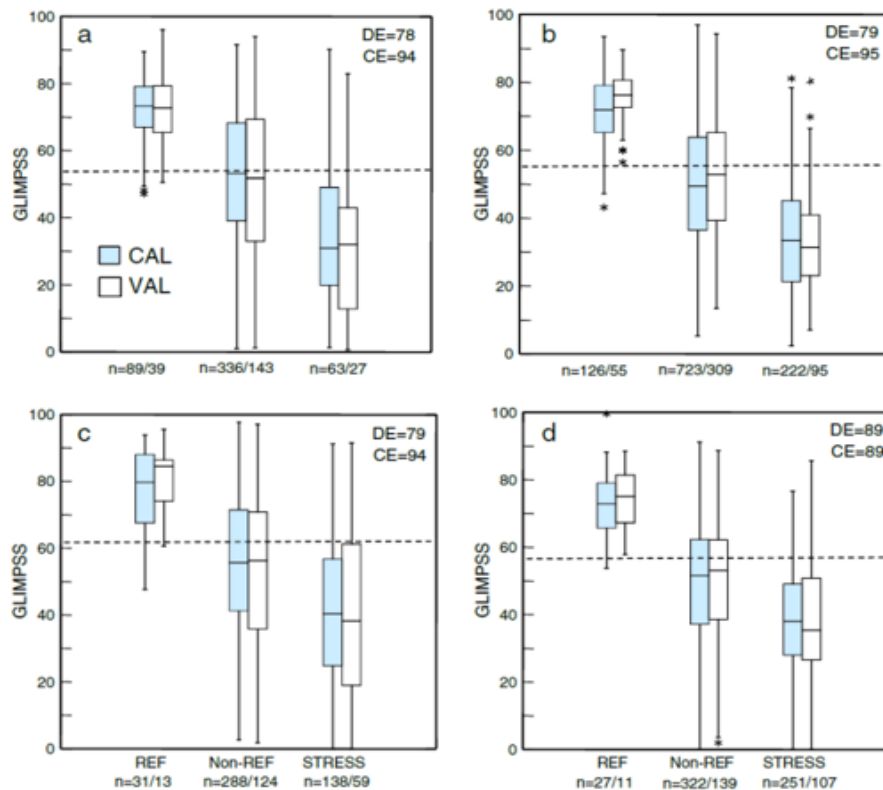


Figure 2. Boxplots of calibration (CAL) and validation (VAL) GLIMPSS scores between REF, non-REF, and STRESS categories in MT Sp (a) MT Su (b), PL Sp (c) and PL Su (d).

Number of sites in each stratum shown as n=CAL/VAL. Percent discrimination efficient (DE) for CAL and percent classification efficiency (CE) for VAL also provided. Dashed line represents approximate 5th centile of

4. WVDEP additional data

While the GLIMPSS has sufficient accuracy and confidence for all seasons without additional data, EPA encourages all states to continually incorporate new information into existing indices as a scientific best practice. EPA commends WVDEP for continuing to collect additional reference site samples and incorporate them into the GLIMPSS since publication of the 2016 303(d) list. Classification efficiencies improved slightly with these data but highlight that increasing the number of reference sites does not drastically affect index performance when the original index has robust performance. EPA still recommends that WVDEP incorporate these additional reference samples into the scoring of samples for bioassessment.

B. A single macroinvertebrate sample is robust and sufficient to base an assessment determination without a zone of uncertainty

Requiring a second sample for a site with WVSCI scores between 50 and 72 to “confirm” impairment is a modification on the same zone of uncertainty that EPA explained in connection with its disapproval of WVDEP’s 2012 303(d) and is statistically invalid.

If the reasoning behind taking a second sample is to account for potential field and method variability, that variability is already accounted for in the index development and single sample assessment process. First, all samples have measurement error. Requiring a second sample does not eliminate or reduce this error. In other words, there is no reason that a

second sample would more accurately reflect conditions than the initial sample. Second, to determine impairment of a test site compared to a regional reference condition (i.e., WVSCI or GLIMPSS score), a single test site score is compared to the reference distribution of index scores (normally composed of single samples from numerous reference sites that are considered natural or near natural). This approach asks whether a single test site (represented by a single observation) is a member of a population of sites (represented by single samples from numerous reference sites). In most cases, the population of reference sites consists of single scores at reference sites and is not an “error free distribution”. In other words, the distribution of reference site scores (and calculation of reference percentile) includes measurement error (due to field and lab methods) and statistical error (the reference site range is only a sample of the reference sites, not the entire population of reference sites). Because the distribution of reference sites already incorporates measurement error, requiring additional samples or incorporating a “zone of uncertainty” to account for measurement error double counts this error. It is not needed.

It should be noted that the estimate of the 5th percentile of the *sample* reference distribution is only an estimate of the true 5th percentile of the *population* reference distribution. To the extent WVDEP still desires to identify a range of scores to account for statistical error (rather than measurement error) in calculating the reference distribution from reference samples, EPA continues to recommend that WVDEP adopt an appropriate statistical method for deriving it (such as the interval/equivalence statistical test (Kilgour et. al. 1998)). EPA staff remain available to provide technical support to WVDEP to ensure WVDEP’s assessment methodology is statistically valid.

WVDEP uses a 5th percentile of reference scores (rather than 10th or 25th percentile like neighboring states) as the threshold for attainment. Use of a lower percentile will identify fewer waters as impaired and minimizes the likelihood of a “false positive” score (i.e., one that incorrectly identifies a sample as impaired when it is actually attaining). In statistical terms, the null hypothesis is that the site is a member of the unimpaired (reference) population. Setting a threshold as a percentile of the reference population sets the acceptable significance of the test (α), or the acceptable type 1 error rate (false positive), as the reference percentile. For example, if the threshold is set at the 5th percentile of the reference distribution (as in WV), then any site below that threshold will be rated “impaired”. Note that, on average, 5% of reference sites in the distribution of reference sites will also be rated impaired – these are the false positives, or type I error. (Type 2 error is the error of not detecting impairment when it exists). Setting $\alpha = 0.05$ means that this “false positive” error rate (about 5%) is acceptable. In other words, the chance that a sample will be identified as impaired when it is actually attaining is approximately 5%. EPA notes the 5th percentile is quite low and assumes that most of the reference sites are in a natural state. Further, unless samples were not collected according to methodology, probability analyses indicate that sites scoring *outside* of the reference site distribution (scoring below a WVSCI 72), have less than 0.5% chance of being identified as impaired when they are truly attaining.

Additionally, biological indices of biotic integrity like WVSCI and GLIMPSS are designed, calibrated, and tested to achieve high levels of accuracy and precision to minimize uncertainty and enable assessment decisions with one macroinvertebrate sample. If precision and accuracy is not sufficient, the biotic index would be adjusted to improve it until

acceptable. Evaluating precision is a common quality assurance requirement of most environmental sampling efforts. As discussed previously, the precision of sample scores can be evaluated by comparing duplicate and revisit samples, and the precision for GLIMPSS and WVSCI reference scores is high, indicating one sample is sufficient to adequately characterize the macroinvertebrate community and stream condition (Pond et al., 2011)¹. (Precision metrics are provided in table 2 above.) The calculated mean coefficient of variation of WVSCI and GLIMPSS is less than duplicate water chemistry samples and comparable to indices used by other states. External peer reviewers (through the scientific publication process) accepted the degree of precision and accuracy achieved by GLIMPSS. Use of the lower 5th percentile threshold, combined with the precision of the WVSCI and GLIMPSS indices, indicates that a single sample is sufficient to determine impairment.

C. Use of genus-level data (GLIMPSS) provides more accurate assessments than family-level data (WVSCI)

GLIMPSS is a next generation index designed to provide higher assessment resolution as compared with the WVDEP's existing family-level WVSCI. WVSCI is a family-level multi-metric index and was developed in coordination with EPA in 2000. Since publication of WVSCI in 2000, available biological data and science have progressed significantly. The number of available reference sites has increased, the state of the science has moved from family-level analysis to genus-level analysis, and WVDEP currently has over 6,300 sites that have been sampled and identified using genus-level taxonomy and have GLIMPSS scores calculated. EPA's National Rivers and Streams Assessment and several neighboring states (KY, OH, PA, MD, TN) use genus-level assessment tools.

At the request of WVDEP, EPA worked with WVDEP to develop GLIMPSS for the State of West Virginia. In developing GLIMPSS, 41 different biological metrics were tested across seasonal and geographic strata, primarily to refine expectation for aquatic life use attainment in WV. GLIMPSS was developed using nearly 400 reference sites as opposed to the 107 reference sites originally used to develop WVSCI. GLIMPSS responds favorably to various stressors, providing better diagnostic capabilities than the WVSCI². Increased accuracy of GLIMPSS compared to WVSCI is partially because GLIMPSS accounts for natural variability (driven by geographic location, seasonality, and waterbody size) when WVSCI does not. Through the development of GLIMPSS, WVDEP and EPA learned that these three natural factors influence benthic macroinvertebrate community composition and must be accounted for when scoring sites. If they are not, difference in an individual site score compared to the reference distribution (used to determine impairment) may be attributed to natural factors instead of anthropogenic or vice versa causing error in assessment decisions (see figure 3). For example, healthy macroinvertebrate communities in streams sampled in the Plateau in the summer are not readily comparable to healthy macroinvertebrate communities in Mountain streams sampled in the spring. WVSCI does not distinguish the two, which results in incorrect identification of impairments and attainments when compared

to GLIMPSS. The seasonal-ecoregion (and stream size) calibration of GLIMPSS minimizes this error.

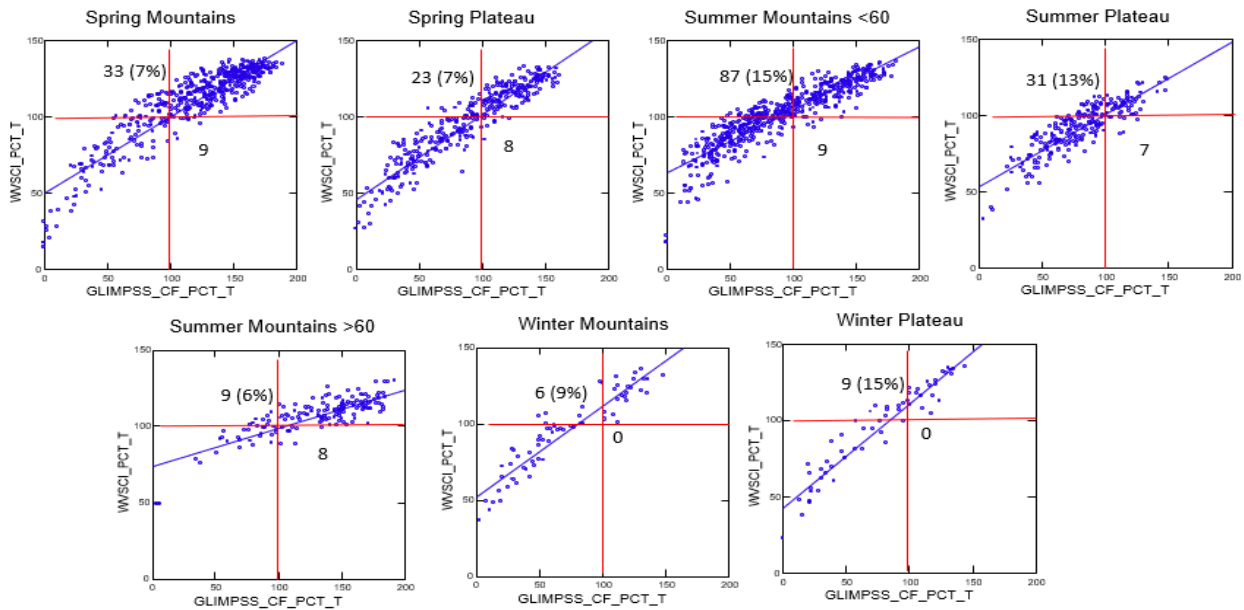


Figure 3: Results of using WWSI instead of GLIMPSS by ecoregion using data from 2013-2017. Samples in the upper left quadrants of each graph are attaining according to WWSI, but impaired according to GLIMPSS. Samples in the lower right quadrant of each graph are impaired according to WWSI, but attaining according to GLIMPSS.

D. Analysis at the family level does not use existing and readily available bioassessment information collected from November to March and information about taxa loss

The available tool to evaluate family level macroinvertebrate data (WWSI) is not calibrated for data collected from November to March, whereas the genus level tool (GLIMPSS) allows use of these data. Accordingly, by not using its genus-level data, WVDEP is not using data from samples collected from November to March.

E. Analysis at the family level does not use information about taxa richness and taxa loss that is available from genus-level data

Because genus-level data provides higher resolution and better reflects sensitivity and taxa loss, WVDEP's use of family level data rather than readily available genus level data also does not use existing and readily available information even for samples collected April to October. Genus level data increases the sensitivity and range of response of biotic indices and improves representativeness.² GLIMPSS allowed for the selection of several additional indicator metrics with larger response ranges which improved tracking of stressors in different seasons and ecoregions. Pond et al. (2008³) reported that the genus Hilsenhoff

³ Pond, G. J., Passmore, M. E., Borsuk, F. A., Reynolds, L., & Rose, C. J. (2008). Downstream effects of mountaintop coal mining: comparing biological conditions using genus- and family-level bioassessment tools. *Journal of the North American Benthological Society*, 27, 717–737.

Biotic Index (HBI) metric was more sensitive than the family HBI metric, the GLIMPSS had wider range of metric responses, and the GLIMPSS had stronger correlations to stressors among WV unmined and coal mined catchments. For those metrics that count taxa richness, the gain in information can be substantial (e.g., loss of several heptageniid mayfly genera at a site would occur before the family was extirpated)². Sample identification at the genus level taxonomy (GLIMPSS) demonstrated loss of entire functional feeding groups (scrapers), a change in composition that might not be identified using WVSCI (Pond, et al. 2014⁴). (GLIMPSS scoring incorporates information on the scraper and shredder guilds, but WVSCI does not.) Loss of an entire functional feeding group indicates ecosystem imbalance.

The use of genus level biological data is supported by several studies, including a recent analysis performed by Dr. Ryan King from Baylor University, whose review of WVDEP data from the Mountain Summer stratum using Threshold Indicator Taxa Analysis (TITAN) demonstrated that the proposed WVSCI thresholds allow significant loss (reduced frequency and abundance) of sensitive genera. For example, 80% of sensitive genera experienced significant losses (reduced frequency and abundance) at a WVSCI threshold of 72. Below a WVSCI of 61, virtually no sensitive genera remained. The same analysis also revealed that at the family level, 65% of sensitive families experienced significant loss below a WVSCI score of 80. Dr. King presented his findings at the 2019 Society for Freshwater Science Annual Meeting and at WVDEP's April 20, 2020 public hearing on a previously proposed bioassessment rule. An electronic copy of Dr. King's presentation was included with EPA's comments to WVDEP in the March, 2021 letter. There can be as many as 24 genera in one family in WV, with each genus having unique stress tolerances, habits, functional feeding groups, and life history. WVDEP does not use information about differences in macroinvertebrate ecology within a family or known seasonal and ecoregional variations in community expectation that could indicate impairment when it uses WVSCI rather than GLIMPSS to assess stream condition.

In closing, EPA appreciates WVDEP's continued dedication and cooperation in evaluating water quality in West Virginia. We look forward to receiving the final 2018-2020-2022 IR submission. If you have any questions regarding this letter, please contact me, or Leah Ettema at 304-234-0245 ettema.leah@epa.gov.

Sincerely,

Gregory Voigt, Chief
Standards & TMDLs Section

Enclosure

⁴ Pond, GJ, ME Passmore, ND Pointon, JK Felbinger, CA Walker, KJG Krock, JB Fulton, and WL Nash. 2014. Long-term impacts on macroinvertebrates downstream of reclaimed mountaintop mining valley fills in central Appalachia. Environmental Management. DOI 10.1007/s00267-014-0319-6.